Nature Reviews Molecular Cell Biology | AOP, published online 1 October 2014; doi:10.1038/nrm3884

ANALYSIS

The structure, function and evolution of proteins that bind DNA and RNA

William H. Hudson and Eric A. Ortlund

Abstract | Proteins that bind both DNA and RNA typify the ability of a single gene product to perform multiple functions. Such DNA- and RNA-binding proteins (DRBPs) have unique functional characteristics that stem from their specific structural features; these developed early in evolution and are widely conserved. Proteins that bind RNA have typically been considered as functionally distinct from proteins that bind DNA and studied independently. This practice is becoming outdated, in partly owing to the discovery of long non-coding RNAs (lncRNAs) that target DNA-binding proteins. Consequently, DRBPs were found to regulate many cellular processes, including transcription, translation, gene silencing, microRNA biogenesis and telomere maintenance.

Proteins that bind DNA or RNA are often considered and studied independently of one another. For example, transcription factors are usually modelled relatively simply: they bind to genomic promoters and control target gene expression by activating or repressing RNA polymerases. Following transcription, RNA-binding proteins modulate protein expression by regulating the stability and translation of mRNAs. However, the consideration of DNA- and RNA-binding functions within proteins as separate entities is becoming outdated. The unappreciated dual DNA- and RNA-binding capacity of a growing body of proteins plays a key part in modulating gene expression, cell survival and homeostasis. Recent studies have demonstrated that many transcription factors are capable of binding diverse types of RNA, which enables them to bind to the mRNA products of transcription to regulate their turnover and to integrate other signals, such as responses to stress[1–7]. Additionally, the prevalence and emerging functions of long non-coding RNAs (lncRNAs) have revealed that non-coding RNAs target many types of proteins through direct interactions[1,8–11].

In this Analysis, we enumerate these DNA- and RNA-binding proteins (DRBPs) and describe their functions, structures and evolution. We first broadly discuss the prevalence of DRBPs within the human genome. We highlight known functions of DRBPs with specific examples of how the simultaneous and serial RNA and DNA interaction allows better gene targeting, finer control of gene expression and integration of metabolic state or stress to modulate protein activity. We discuss the structural features of DRBPs that enable dual nucleic acid specificity, focusing on the limited number of solved structures that allow direct comparison of a DRBP complexed with either DNA or RNA. Finally, we discuss the evolution of dual DNA- and RNA-binding domains within DRBPs, including ancient domains, for which dual DNA and RNA binding conferred a selective advantage, and more modern domains, which have recently been targeted by rapidly evolving lncRNAs.

Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30033, USA.
Correspondence to E.A.O.
e-mail: eortlun@emory.edu
doi:10.1038/nrm3884
Published online
1 October 2014

Defining DRBPs

Defining the subset of human proteins that bind both DNA and RNA is a difficult task. Using gene ontology searches, only 64 human protein-coding genes in the QuickGO gene ontology database[12] (European Bioinformatics Institute) are identified as having direct and specific experimental evidence for both RNA binding (GO:0003723) and DNA binding (GO:0003677) (FIG. 1a). The PROTEOME database (BioBase) returns 122 such proteins, although direct evidence is lacking for many of them.

An alternative approach involves combining evidence from studies that have separately attempted to catalogue all human proteins that bind DNA or RNA (FIG. 1b). A study using protein microarrays and bioinformatic approaches identified >4,000 human proteins that directly interact with double-stranded DNA (dsDNA) *in vitro*[13]. Gene ontology analysis of these proteins reveals that the group is highly enriched for the term "RNA binding" ($P < 1 \times 10^{-40}$), indicating that RNA binding may be a common feature of DNA-binding proteins (FIG. 1c). Among these dsDNA-binding proteins, the ontology term "dsRNA binding" is much more represented than "ssRNA binding" (single-stranded RNA binding).

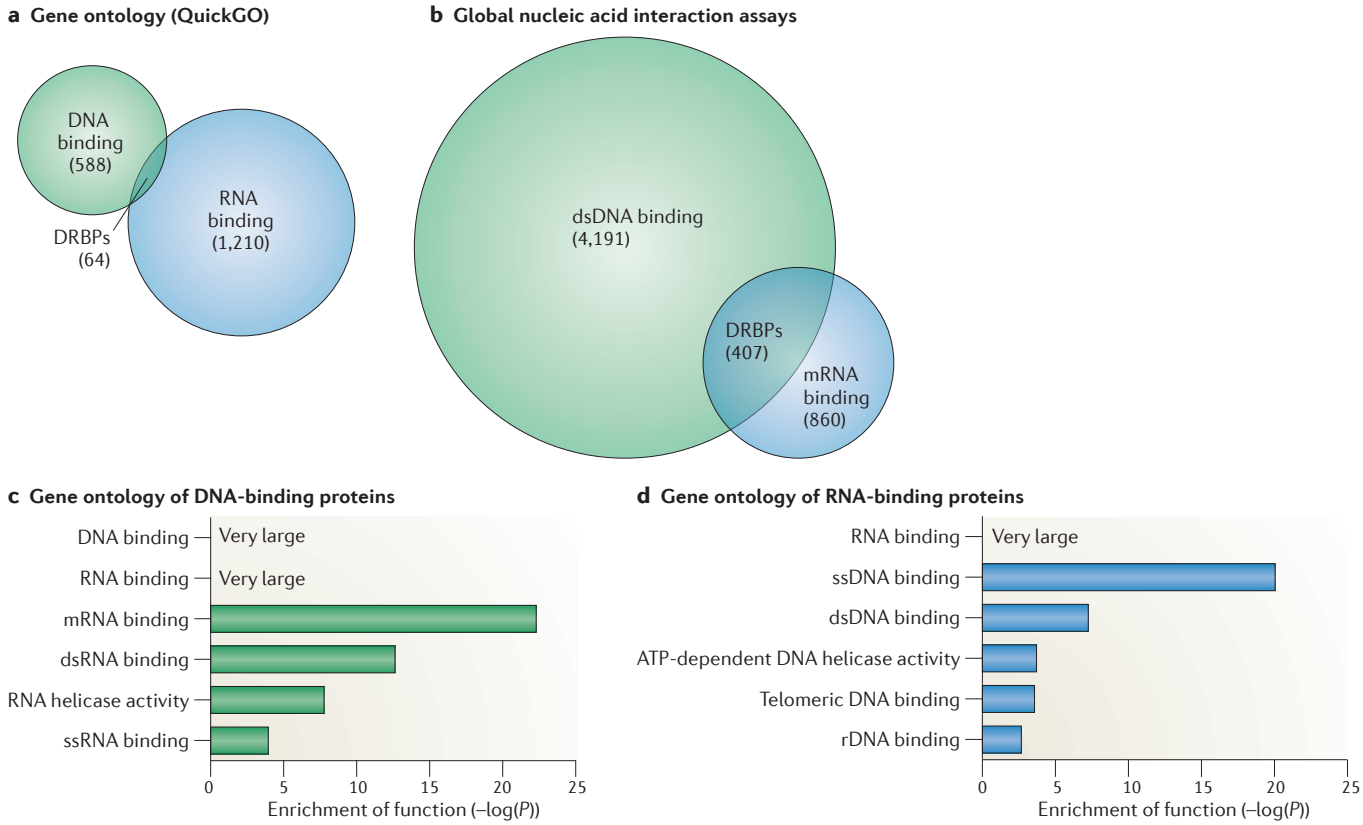

Figure 1 | **Defining human DRBPs.** **a** | Venn diagram shows DNA-binding proteins and RNA-binding proteins in the QuickGO database supported by low-throughput experimental evidence (as of July 2014)[12]. The overlap of these two sets represents human DNA- and RNA-binding proteins (DRBPs), which consists of 64 proteins. **b** | Venn diagram shows DNA-binding proteins and RNA-binding proteins identified in high-throughput studies defining the human mRNA and double-stranded DNA (dsDNA) interactomes[13,14]. There are 407 proteins found in both studies, indicating that they may bind both mRNA and dsDNA. In parts **a** and **b**, circles are drawn to scale. **c** | Molecular function gene ontology analysis reveals that RNA binding is a potentially major function of the dsDNA-binding proteins identified in REF. 13. **d** | Gene ontology analysis reveals that DNA binding is potentially a major function of the mRNA-binding proteins identified in REF. 14. In parts **c** and **d**, only selected molecular function attributes are shown for brevity. *P* values in parts **c** and **d** indicate the probability that the over-representation of the stated ontology term in the selected 407 genes compared with all human genes is due to chance. These were calculated in the TRANSFAC + PROTEOME database (BioBase) using the hypergeometric distribution; "very large" indicates a *P* value of $<1\times10^{-40}$ (–log(*P*) >40). rDNA, ribosomal DNA; ssDNA, single-stranded DNA.

Another study used a crosslinking- and mass spectrometry-based approach to identify 860 mRNA-binding proteins from HeLa cells termed the mRNA interactome[14]. Functional analysis of these proteins again indicates that dual nucleic acid binding is a widespread phenomenon (FIG. 1d), as they are significantly enriched for both ssDNA binding and dsDNA binding ($P=8.9\times10^{-21}$ and $P=5.5\times10^{-8}$, respectively). Notably, of the 860 proteins identified as mRNA binding, 407 (47.3%) were independently characterized as dsDNA binding in REF. 2. Together, the two studies indicate that DRBPs are widespread, perhaps constituting 2% of the human proteome (407 of ~20,300 proteins, FIG. 1b). This number would probably increase if the studies included proteins that are expressed in other cell types, proteins that require ligand binding-dependent signals for nucleic acid binding, or proteins that bind other types of DNA or RNA.

We note that many of the proteins identified in REFS 13,14 as DNA and/or RNA binding lack corroborating evidence from other studies, and these findings should thus be interpreted with caution. For example, many identified proteins, such as polymerase subunits, may bind nucleic acid-bound proteins without binding DNA and/or RNA directly. Additionally, many proteins that bind DNA or RNA *in vitro* may not bind them *in vivo*. However, the two studies provide a reasonable estimate of potential human DRBPs owing to their wide coverage of the human proteome, and we discuss below examples in which the demonstration of protein–nucleic acid binding *in vitro* has preceded the discovery of such binding *in vivo*, sometimes by decades.

In Supplementary information S1 (table), we provide a detailed list of 149 human DRBPs, with comments on their nucleic acid-binding properties, structures and functions. These proteins were selected based

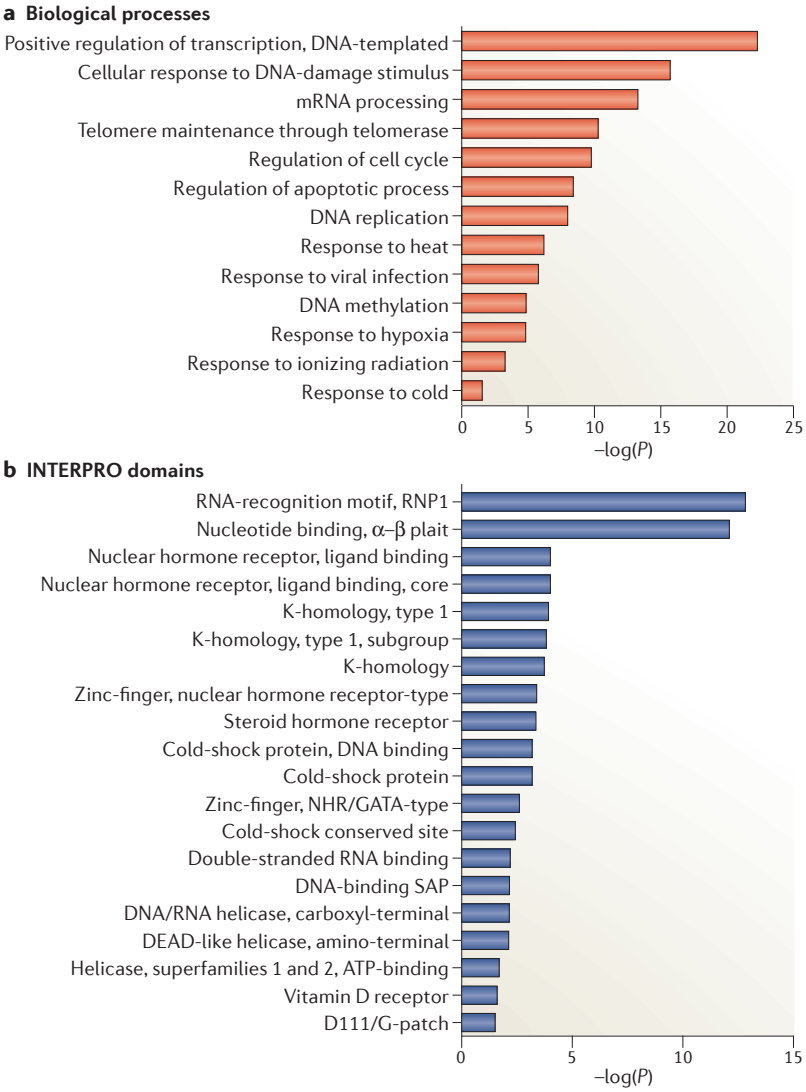

Figure 2 | **Functional and structural properties of DRBPs.** The 149 DNA- and RNA-binding proteins (DRBPs; see Supplementary information S1 (table)) were subjected to gene ontology enrichment of biological process (PROTEOME database; BioBase) and to INTERPRO domain enrichment (DAVID ontology[77,78]) in order to explore the biological functions of and protein domains commonly found in DRBPs. **a** | In gene ontology analysis, biological processes such as transcriptional regulation and mRNA processing are expectedly prominent terms found to be enriched in DRBPs. However, unexpected functions are also enriched, including response to many cellular stresses (such as heat, viral infection and radiation). For brevity, only selected functions are shown. **b** | In domain enrichment analysis, all domains enriched in the set of 149 DRBPs that have *P* values $\leq 10^{-3}$ are shown. *P* values in parts **a** and **b** indicate the probability that the over-representation of the stated term in the 149 DRBPs compared with all human genes is due to chance. RNP1, ribonucleoprotein 1

on experimental evidence demonstrating their ability to bind directly to both DNA and RNA, generally obtained from studies using more traditional experimental approaches than the high-throughput studies[13,14] discussed above. Although many of the proteins in Supplementary information S1 (table) have only been shown to bind DNA and/or RNA *in vitro*, the remainder of this Analysis focuses on selected human DRBPs with known cellular roles.

Functions of DRBPs

We carried out gene ontology and domain enrichment analyses (FIG. 2) to illuminate the main biological functions of our list of human DRBPs (see Supplementary information S1 (table)). The gene ontology analysis revealed expected biological processes, such as transcriptional regulation, mRNA processing and DNA replication. However, several surprising functions are also implicated, including the DNA-damage response, apoptosis and responses to extreme temperatures (FIG. 2a).

Ultimately, DRBP functions are governed by their inherent structural and biochemical properties. One can envision DRBPs capitalizing on both RNA and DNA binding in numerous ways; for example, a transcription factor that binds DNA and RNA may interact orthogonally with RNAs that compete with DNA binding to repress transcription, or simultaneously with a promoter and an RNA co-activator to upregulate transcription. The following section focuses on DRBPs that bind DNA and RNA competitively (FIG. 3a).

Binding DNA or decoy RNAs. The role of certain lncRNAs as decoys of genomic DNA is illustrated by the reduction in promoter occupancy by transcription factors, typically measured by chromatin immunoprecipitation (ChIP), in response to the overexpression of competing decoy RNAs. The glucocorticoid receptor (GCR), a steroid hormone receptor, is a classic example of a ligand-activated transcription factor (reviewed in REF. 15). In its inactivated state, the GCR is kept in the cytoplasm by chaperone proteins. Upon ligand binding, the GCR translocates to the nucleus, where it can bind to the promoters, and regulate the transcription, of hundreds of genes[16]. Given the anti-inflammatory role of the GCR, much effort has been put into developing modulators of GCR-driven transcription[17]. Recently, the lncRNA growth arrest-specific 5 (*GAS5*) was found to inhibit the transcriptional activity of the GCR by competing directly with DNA for protein binding *in vitro* and in cells[1]; overexpression of *GAS5* leads to a decrease in ChIP-detected GCR occupancy at its target promoters, as well as a decrease in the mRNA levels of glucocorticoid-activated genes[1,18]. As cellular *GAS5* levels are regulated by nonsense-mediated decay[18] in response to serum starvation and other stressors[1], the transcriptional activity of the GCR is tuned by titrating the levels of *GAS5* against the fixed number of genomic GCR-binding sites in response to cellular stress. Three closely related steroid receptors that share the DNA specificity of the GCR — the androgen, progesterone and mineralocorticoid receptors — are also susceptible to *GAS5*-mediated transcriptional repression[1]. Although steroid receptors have traditionally been thought of as DNA-binding proteins, the affinity of the GCR for RNA and DNA is similar, as measured *in vitro* by glutathione *S*-transferase pulldown assays and fluorescence-based competition assays[1]. The most distantly related member of the steroid receptor family, the oestrogen receptor, does not share the DNA specificity of the GCR and is not susceptible to *GAS5*-mediated transcriptional repression, indicating that the binding of steroid receptors to RNA is sequence specific[1].

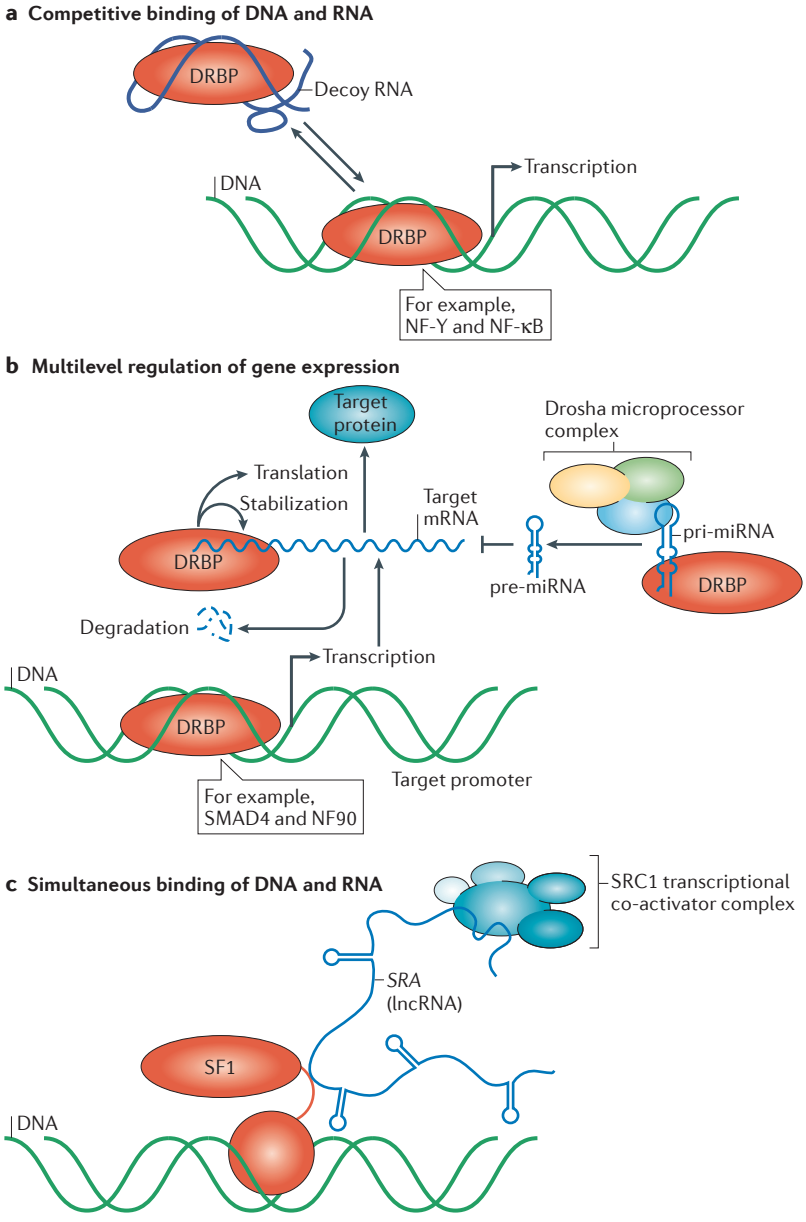

Figure 3 | **Three archetypes of DRBP function.** **a** | RNA can compete with DNA for binding to DNA- and RNA-binding proteins (DRBPs), typically at the same protein interface. In the case of transcription factors, this can reduce promoter occupancy and the transcription of target genes. **b** | DRBPs can regulate gene expression at multiple levels. In addition to binding to the promoters of genes to regulate their transcription, DRBPs can also affect microRNA (miRNA) processing, as well as mRNA stability and translation. **c** | DRBPs can bind DNA and RNA simultaneously, allowing the RNA to function as a scaffold to recruit other proteins to a specific DNA locus. Shown here is steroidogenic factor 1 (SF1) binding to the long non-coding RNA (lncRNA) steroid receptor RNA activator (*SRA*) to recruit the SRC1 (also known as NCOA1) transcriptional complex in a ligand-independent manner. NF, nuclear factor; NF90, NF of activated T cells 90 kDa; pre-miRNA, precursor miRNA; pri-miRNA, primary miRNA.

Rossmann fold
A common protein-folding pattern that contains the topology of a β-α-β fold. It is found in many nucleotide-binding proteins.

Additional examples of pairs of transcription factors and decoy RNAs are nuclear factor-Y (NF-Y), which binds the lncRNA P21-associated ncRNA DNA-damage activated (*PANDA*)[8], and NF-κB, which binds the mouse pseudogene-derived RNA *Lethe*[9]. The dual nucleic acid-binding activity of NF-κB was demonstrated *in vitro* many years before the discovery of an endogenous RNA target[10], suggesting that transcription factors that are known DRBPs *in vitro* may have endogenous RNA targets awaiting discovery; an example of such DRBPs is the acute myeloid leukaemia 1 (AML1; also known as RUNX1) protein[19]. Although structural information on the interaction of human proteins with their decoy RNAs is lacking, a recent study demonstrated an elaborate mechanism of an analogous bacterial system: the sequestration of ribosomal RNA small subunit methyltransferase E (RsmE) by the non-coding RNA *RsmZ*[11]. Competitive DNA and RNA binding is a feature not only of transcription factors but also of nucleic acid-modifying enzymes, such as DNA methyltransferases. In humans, DNA methylation is initiated by DNA (cytosine-5)-methyltransferase 3A (DNMT3A) and DNMT3B; DNMT1 maintains this methylation by binding to hemimethylated DNA after replication (reviewed in REF. 20). RNA binding can inhibit the DNA-binding and methylation activity of both DNMT3A[21] and DNMT1 (REF. 22). *In vitro*, DNMT1 binds RNA with a higher affinity than DNA, as shown in electrophoretic mobility shift assays (EMSA)[21]. In the case of DNMT1, and probably in that of DNMT3A, RNAs bind to the catalytic domain of the methyltransferase to inhibit DNA methylation[21,22].

It is notable that several metabolic enzymes — such as the glycolytic enzymes lactate dehydrogenase[23–25], glyceraldehyde-3-phosphate dehydrogenase (GAPDH)[26–28] and α-enolase (ENO1)[14,29,30] — are DRBPs with competitive DNA- and RNA-binding capacities. In the case of GAPDH, DNA and RNA compete for binding of the cofactor NAD^+ to the enzyme[26,28], suggesting that Rossmann fold-containing proteins such as GAPDH may be sensitive to cellular DNA and/or RNA levels. ENO1 binds RNA as a monomer[30], which inhibits the formation of the catalytically active protein dimer[31,32]. NAD^+-specific isocitrate dehydrogenase, which converts isocitrate to α-ketoglutarate, is allosterically inhibited by the 5′ untranslated regions of yeast mitochondrial mRNA[33]. Binding of RNA and DNA to metabolic enzymes indicates that nucleic acids can regulate the function of proteins other than transcription factors to modulate cellular metabolism[34].

DRBPs that regulate gene expression at multiple levels. Approximately half of the DRBPs we identified in our analysis are transcription factors. As discussed above, some such proteins have been shown to be the targets of decoy RNAs. By contrast, several others bind both the DNA and the mRNA of their target genes (FIG. 3b). Regulating genes at both the DNA and the RNA levels allows powerful, combinatorial control over protein expression and may enable DRBPs to generate both immediate effects (through regulating RNA turnover) and long-lasting effects (through regulating transcription).

When activated, the GCR can promote the transcription of anti-inflammatory genes[35] and repress the transcription of pro-inflammatory genes[36,37]. Agonist-bound GCR destabilizes the mRNA of pro-inflammatory genes such as the chemokine (C-C motif) ligand 2 (*CCL2*; also

known as *MCP1*) gene through direct RNA binding, perhaps by the recruitment of ribonucleases[38]. The presence of a GCR-binding motif in many immunogenic mRNAs has been confirmed using RNA immunoprecipitation (RNA-IP) and suggests that the GCR can accelerate the decay of many mRNAs, broadening its known role in the anti-inflammatory reponse[2]. Given that the GCR also binds directly to pro-inflammatory transcription factors, such as adaptor protein complex 1 and NF-κB[39,40], it seems that the GCR uses its diverse DNA-, RNA- and protein-binding capacities to regulate inflammatory genes at the transcriptional and post-transcriptional level.

Transcription factors can also regulate gene expression post-transcriptionally through the regulation of microRNA (miRNA) biogenesis. miRNAs are small RNAs that facilitate gene silencing through sequence-specific pairing to target mRNAs and recruitment of these to the RNA-induced silencing complex (RISC; reviewed in REF. 41). Several transcription factors have been shown to regulate Drosha-mediated primary miRNA (pri-miRNA) processing, a key step in the biogenesis of functional miRNAs[42]. SMAD proteins, which are transducers of transforming growth factor-β (TGFβ) signalling, activate transcription by forming a DNA-binding heterodimer (reviewed in REF. 43). SMAD proteins also increase the levels of several miRNAs, including miR-21 (REF. 44), which has important roles in development and immunity[45]. Surprisingly, the increase in miR-21 levels is due not to increased transcription of pri-miR-21 but to increased Drosha-mediated processing of pri-miR-21 to the precursor miRNA (pre-miRNA) mir-21 (REF. 44). Bioinformatic analysis identified a conserved RNA motif in TGFβ-regulated miRNAs, which was shown by RNA-IP and EMSA to bind directly to the MAD homology 1 domain of SMAD to mediate Drosha processing[3]. Interestingly, the RNA sequence motif that is bound by SMAD4 and that mediates the regulation of miRNA expression post-transcriptionally is identical to the DNA sequences that are bound by SMAD4 and that mediate regulation of gene expression transcriptionally[3].

NF of activated T cells 90 kDa (NF90; also known as ILF3) is a particularly versatile DRBP that, along with its partner NF45 (also known as ILF2), has important roles in T cell activation[46]. Through the direct binding of DNA, mRNA and miRNA, NF90 controls transcription[4,5], regulates mRNA turnover and translation[47,48], and affects miRNA processing[49], respectively. These functions assist in its role in T cell activation: NF90 upregulates the mRNA levels of interleukin-2 (*IL-2*), a critical cytokine in T cell development[50], by binding its promoter to activate its transcription and by stabilizing the *IL-2* mRNA through direct binding to its 3′ untranslated region, as was found by EMSA and ribonucleoprotein (RNP) immunoprecipitation analysis[5,48]. Additionally, using *in vitro* pri-miRNA-processing assays and RNA-IP, NF90 in complex with NF45 was shown to inhibit the processing of the pri-miRNA pri-let-7a by binding it directly[49]. let-7a represses IL-6, a cytokine that is critical for T cell survival and proliferation[51], which may link inflammation to cancer[52], and let-7 downregulation following NF90 upregulation reduces survival in several cancer types[53,54]. In summary, these examples illustrate that DRBPs can use both transcriptional and post-transcriptional mechanisms to serve as potent controllers of gene expression.

Simultaneous binding of DNA and RNA. In contrast to DRBPs that target DNA or RNA serially to serve different or related functions, another class of DRBPs binds RNA and DNA simultaneously to perform a single function (FIG. 3c). Generally, transcription factors not only require DNA binding to target promoters but also bind to co-repressors or co-activators to affect transcriptional regulation. There are several examples of RNA molecules acting as co-activators by simultaneously binding DNA and various transcription factors. The lncRNA rhabdomyosarcoma 2-associated transcript (*RMST*), in particular, is required for binding of neurogenic gene promoters and subsequent upregulation by SOX2 (REF. 55), a transcription factor with important roles in development, pluripotency and cell fate[56]. RNA-IP and RNA pulldown experiments showed that *RMST* interacts directly with SOX2 (REFS 55,57), and DNA occupancy of SOX2 measured by ChIP followed by sequencing (ChIP–seq) was reduced following *RMST* depletion[55,57]. The lncRNA *EVF2* is a transcriptional co-activator of DLX2 (REF. 58) and recruits methyl-CpG-binding protein 2 (MECP2) to intergenic enhancers[59]. A direct interaction between DLX2 and *EVF2* has been demonstrated by the immunoprecipitation of DLX2 followed by reverse transcription PCR of the *EVF2* lncRNA[58], and MECP2 also has previously been shown to bind RNA[60]. It should be noted that RNA-mediated recruitment of a protein to a particular DNA locus might not require direct binding of both DNA and RNA by the protein, as lncRNAs could recruit transcription factors to a particular DNA locus to which the lncRNA is bound. Dual nucleic acid recognition also facilitates targeted gene repression through RNA-guided DNA methylation. This phenomenon was first discovered in plants[61], and some mammalian RNA guides of DNA methylation have since been found[62,63], although their mechanisms of action are less clear. In mice, DNMT3A forms a complex with *Tsix* RNA to promote methylation of the X-inactive-specific transcript (*Xist*) promoter[64].

Several nuclear receptors — including steroidogenic factor 1 (SF1), DAX1 (also known as NR0B1) and thyroid receptor-α (TRα) — bind simultaneously to both gene promoters[65,66] and the RNA co-activator *SRA* (steroid receptor RNA activator) to modulate transcriptional activation[6,7] (FIG. 3c). Using pulldown experiments, SF1 and TRα have been shown to bind *SRA* through their hinges, which are flexible, disordered regions that connect their DNA- and ligand-binding domains[6,67]. Knockdown of *SRA* decreases the interaction of SF1 with protein transcriptional activators and the transcription of SF1-regulated genes[6]. Several other nuclear receptors associate with, but lack direct evidence for direct binding to, *SRA*, including the androgen, progesterone and oestrogen receptors, as well as retinoic acid receptor-α (RARα), which may bind *SRA* and its target gene promoters simultaneously[7,68–70].

Drosha
A nuclease of the RNase III family that, as part of a complex, cleaves primary microRNA (pri-miRNA) transcripts into precursor miRNAs (pre-miRNAs) in the nucleus.

Primary miRNA
(Pri-miRNA). The primary transcripts of microRNA-encoding genes, which are stem–loop structures processed by the Drosha complex into precursor miRNAs (pre-miRNAs). Typical pri-miRNAs are hundreds of base pairs long and can contain several pre-miRNAs.

Precursor miRNA
(Pre-miRNA). The product of Drosha-mediated cleavage of primary microRNAs (pri-miRNAs). Stem–loop-structured pre-miRNAs are exported to the cytoplasm and further processed by Dicer to fully mature miRNAs.

MAD homology 1 domain
An evolutionarily conserved domain found in SMAD proteins. It contains four α-helices, six short β-strands and five loops, and it recognizes specific DNA sequences.

Crosslinking immunoprecipitation has demonstrated that RARα can bind to and regulate the translation of target mRNAs through a unique RNA-binding motif at its carboxyl terminus[71].

Another example of simultaneous DNA and RNA binding that is required for DRBP function is the role of telomeric repeat-binding factor 2 (TRF2; also known as TERF2) at telomeres. Deletion of TRF2 leads to an arrest in cell division caused by the formation of chromosome end-fusions[72]. Crystal structures have revealed that TRF2 binds to telomeric DNA in a sequence-specific manner through a C-terminal DNA-binding domain, which resembles a homeodomain[73]. Part of the role of TRF2 at the telomere includes the recruitment, through its positively charged amino-terminal GAR domain, of the origin-recognition complex (ORC, which is a collection of proteins that serves as a scaffold for DNA replication factors, among other functions[74]) so that it can assist in the maintenance of telomere structure[75]. Using biotinylated RNA pulldown experiments, RNA-IP and EMSAs, the GAR domain responsible for ORC recruitment was later shown to bind telomere repeat-encoding RNA (*TERRA*)[76]. Depletion of *TERRA* hampers ORC recruitment to the telomere without affecting TRF2 binding to the telomere itself, suggesting a model in which TRF2 serves as a mediator between telomere DNA and *TERRA*, which in turn recruits factors required for telomere maintenance[76].

Structural characteristics of DRBPs

For a protein such as TRF2 to coordinate telomeric DNA binding and recruit protein complexes by binding of RNA, it must have multiple nucleic acid-binding motifs. Some DRBPs, such as the GCR and NF-κB, have maintained domains that are capable of binding both DNA and RNA, which allows decoy RNAs to evolve and compete with DNA for protein binding. We analyse below the prevalence of structural domains in DRBPs and discuss examples of DRBPs that bind both single-stranded and double-stranded nucleic acids.

DRBP domains that enable DNA and RNA interactions. We carried out InterPro domain enrichment analysis by DAVID[77,78] on our 149 DRBPs to identify domains enriched in proteins that bind both DNA and RNA (FIG. 2b). The RNA-recognition motif (RRM; also known as the RNP domain or RNA-binding domain) was the most highly enriched domain in DRBPs ($P = 2 \times 10^{-26}$). The RRM is an abundant, short (~100-amino-acid long) domain that generally recognizes ssRNA and is often present in proteins with other domains, such as zinc-finger domains, WW domains or additional RRMs[79]. Such multidomain DRBPs may bind RNA and DNA simultaneously through separate domains, as occurs with heterogenous nuclear RNA A1, which contains two RRMs[80]. Single RRM-containing proteins are also capable of binding both DNA and RNA, and this function is present, for example, in RNA-binding motif 3 (RBM3), TBP-associated factor 15 (TAF15), and TAR DNA-binding protein 43 (TDP43; also known as TARDBP)[81–83] (see Supplementary information S1 (table)). Such bivalent domains may not have the same sequence specificity when binding DNA and RNA, highlighting the complexity of recognizing nucleotide bases in a sequence-dependent manner.

Nuclear receptor domains are also enriched in DRBPs: RARα binds mRNA through a unique C-terminal domain[71], SF1 binds the RNA co-activator *SRA* through its hinge and a unique Ftz-F1 domain, and TRα binds *SRA* through its hinge[6,67]. The majority of nuclear receptors have two highly conserved Cys_4 zinc-fingers through which they bind DNA, and some nuclear receptors, such as the GCR, can also bind RNA through these domains[1]. Other types of zinc-fingers are also enriched in DRBPs, such as the RAN-binding protein 2 (RANBP2) type. Other notably enriched domains in DRBPs are the K-homology domain, the dsRNA-binding domain (dsRBD), the cold-shock domain (CSD) and various helicase domains. Each of these domains is capable of binding DNA and RNA, and we focus below on the structural mechanisms underlying this dual specificity.

General properties of DRBPs. There are only two chemical differences between RNA and DNA. First, RNA (but not DNA) has a 2′ hydroxyl (2′OH) group on the ribose sugar, which allows an additional hydrogen bond to be formed and a greater diversity of secondary structures than is possible in DNA. Second, RNA contains uracil rather than thymine, as in DNA; uracil lacks a methyl group at the C5 position. A comparative analysis of known protein–nucleic acid structures revealed that the recognition of DNA occurs largely through electrostatics and direct base–protein interactions. Conversely, RNA recognition by proteins mainly depends on shape complementarity and interaction with the 2′OH group[81]. Given these general differences, one could expect that, during evolution, highly selective protein interfaces would be generated that are optimized for either RNA or DNA, with minimal cross-binding. However, the most energetically favourable associations between proteins and nucleic acids relay on hydrophobic and charge–charge interactions. These interactions are less constrained than interactions with the sugar backbone or with the nucleotide base edge, which is capable of highly-specific Watson–Crick base pairing. Thus, DRBP domains that competitively bind DNA and RNA probably rely on the less specific hydrophobic and charge–charge interactions. For example, ssRNA-binding proteins are more likely to form hydrogen bonds with bases rather than with the phosphate–sugar backbone, compared to those that recognize folded RNA, such as ribosomal proteins and tRNA synthetases[82]. Because ssRNA-binding proteins do not rely heavily on sugar recognition, they are more likely to also bind DNA. This may explain why the RRM is the most enriched domain in the DRBPs included in our analysis[82] (FIG. 2b).

The RRM. The RRM is an extremely versatile domain that is capable of binding (mainly single-stranded) RNA and DNA, as well as proteins[79]. RRMs preferentially interact with nucleic acid bases rather than with the phosphate–sugar backbone. The structural nature of ssRNA and ssDNA allows much easier access to the exposed aromatic base faces, as opposed to hydrogen bonding to the

Homeodomain
A protein structural domain that binds DNA and RNA, and that is found most commonly in transcription factors. It consists of a helix–turn–helix structure of 60 amino acids.

GAR domain
A Gly- and Arg-rich motif that adopts a repeated β-turn structure. It is found most commonly in proteins that bind RNA.

Zinc-finger domains
Structural motifs that are characterized by the coordination of one or more zinc ions. There are several different zinc-finger motifs, and each displays a different binding mode and structure.

WW domains
Small protein motifs of 40 amino acids that mediate specific protein–protein interactions with short Pro-rich or Pro-containing motifs.

Ftz-F1 domain
A protein domain first identified in the *Fushi Tarazu* factor 1 (FTZ-F1) nuclear receptor. It contains an evolutionarily conserved LXXLL motif that recognizes other LXXLL-related motifs.

K-homology domain
A conserved protein domain that interacts with both RNA and DNA through a binding cleft formed between two α-helices, two β-sheets and the GXXG loop.

Cold-shock domain
(CSD). A small (~70 kDa) domain with high similarity to the ribonucleoprotein 1 RNA-binding motif. This domain is found in DNA- and RNA-binding proteins in bacteria, archaea and eukaryotes.

Interfaces
The solvent-accessible portions of a protein that are capable of binding a ligand — including DNA and RNA — in a competitive or non-competitive manner.

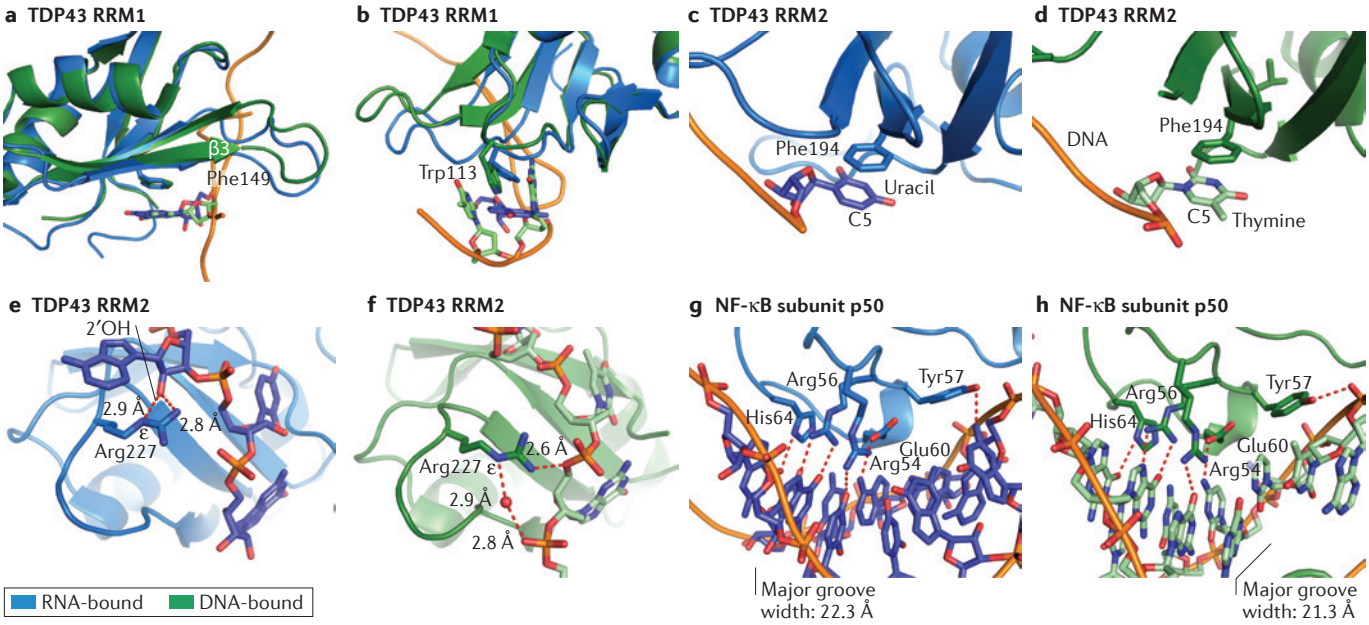

Figure 4 | **The structural basis for dual DNA and RNA recognition by TDP43 and by the NF-κB subunit p50.** Protein–RNA structures are shown in blue and protein–DNA structures in green, with protein in the darker shade. π-stacking interactions play a prominent part in both the single-stranded DNA (ssDNA)- and the ssRNA-binding activities of TAR DNA-binding protein 43 (TDP43). **a**,**b** | Phe149 (part **a**) and Trp113 (part **b**) within the first RNA-recognition motif (RRM1) of TDP43 stack with both RNA and DNA bases. **c**,**d** | In the second RRM (RRM2) of TDP43, Phe194 is capable of recognizing both uracil in RNA (part **c**) and thymine in DNA (part **d**); the additional methyl group at C5 in thymine does not contribute to nucleic acid specificity. **e** | When bound to RNA, both the terminal amine and ε-nitrogen of Arg227 in RRM2 of TDP43 contact a 2′ hydroxyl (2′OH) group on the RNA backbone. **f** | By contrast, these same groups can also make contacts with the DNA backbone, both directly and through water-mediated hydrogen bonding. **g**,**h** | The p50 subunit of nuclear factor-κB (NF-κB) makes strikingly similar base-specific contacts when bound to an RNA aptamer (part **g**) or to double-stranded DNA (part **h**). This is due in large part to the similar secondary structure and chemical moieties presented by the RNA and DNA. Major groove width was calculated with 3DNA software using phosphate–phosphate distances[92].

base edges that occur frequently with double-stranded nucleic acid-binding DRBPs. Additionally, stacking interactions with the faces of bases are more energetically favourable than recognition of the nucleotide edge. Therefore, stacking interactions between aromatic protein side chains and nucleic acid bases are often observed in single-stranded nucleic acid-binding proteins.

TDP43 is a DRBP that has important roles in mRNA splicing and miRNA biogenesis[83,84]. It contains two RRMs, which are separated by a short loop and are both capable of binding DNA and RNA. Crystal structures of the TDP43 RRMs in complex with DNA and RNA have been reported[85,86] (see also the currently unpublished structure of TDP43 RRMs), making TDP43 an excellent case study for dual DNA and RNA recognition by RRMs. The DNA- and RNA-bound structures of TDP43 reveal nearly identical modes of nucleic acid recognition. Aromatic side chains, such as Phe149 within the first RRM (RRM1), form stacking interactions with DNA or RNA bases (FIG. 4a). Trp113, which is part of the more flexible loop 1, is able to shift conformations and base-stack slightly differently when bound to different nucleic acid sequences (FIG. 4b), whereas Phe149 in the rigid β_3 sheet of the RRM1 fold makes similar interactions with DNA and RNA (FIG. 4a). Relying on the more energetically favourable π-stacking interactions through the planar face of the DNA and RNA bases results in less specificity than that gained from hydrogen bonding with the base edge. Uracil and thymine interact with Phe194 of the second RRM (RRM2) in the RNA- and DNA-bound structures, respectively (FIG. 4c,d). Despite the additional methyl moiety at position C5 in the DNA, no TDP43 residues recognize the edge of the nucleotide to interact or clash with the additional carbon (FIG. 4d). Thus, one of the chemical differences between DNA and RNA, the use of uracil in RNA, plays no part in nucleic acid discrimination in this example.

By contrast, the RRM2 of TDP43 does make RNA-specific contacts with the 2′OH group. The majority of protein–2′OH group interactions are mediated through protein side chains[81], and both the Lys263 and Arg227 (FIG. 4e) side chains in RRM2 contact a 2′OH group when bound to RNA. However, when TDP43 is bound to DNA, these same protein side chains contact the DNA backbone phosphates (FIG. 4f), demonstrating that amino acids are capable of reorienting to allow distinct types of interactions to support RNA and DNA binding. Nevertheless, DNA binding is not a general property of all RRMs. For example, the RRM of poly(A)-binding protein relies on many RNA-specific 2′OH contacts for RNA interaction, and binding to DNA may be at low affinity, if detectable at all[87].

π-stacking interactions
Non-covalent interactions between two aromatic molecules owing to the attractive force originating from the opposing electrostatic potentials between two adjacent aromatic amino acid residues.

DRBPs that recognize double-stranded nucleic acids. Crystal structures of protein–dsRNA complexes are less common than their single-stranded counterparts, but there are some examples that are instructive for dual nucleic acid recognition. NF-κB is a central transcription factor of immune signalling and is formed of homodimers or heterodimers of Rel family proteins, such as p50 (also known as NFKB1) or p65 (also known as RELA)[88]. High-affinity aptamers have been developed for both the p50 subunit and the p65 subunit of NF-κB, with an affinity of RNA binding that approaches that of the transcription factor's affinity for native DNA response elements[10,89]. DNA with an identical sequence to the p50-targeting aptamer will not bind p50 (REF. 10); p50 is therefore another DRBP that binds RNA and DNA in a sequence-specific manner, with different sequence specificities for each.

Crystal structures of p50 bound to DNA and to RNA reveal that both bind at the same surface of the p50 immunoglobulin-like domain[90,91]. Although p50 binds to RNA as a monomer and to DNA as a dimer, similar networks of base-specific interactions occur between protein and nucleic acids in each structure (FIG. 4g,h). Not only do the DNA- and RNA-contacting residues of p50 maintain an equivalent position, but both DNA and RNA also present similar interfaces for p50 recognition in charge distribution and in secondary structure[92] (FIG. 4g,h). This is a seminal, structurally confirmed example of 'DNA mimicry' by RNA to bind to a transcription factor and, although the RNA in this case was artificial, DNA mimicry has been hypothesized to play a part in the endogenous regulation of several transcription factors[1,93].

Structures have also been solved of the DRBP dsRNA-specific adenosine deaminase 1 (ADAR1), which binds both double-stranded Z-DNA and double-stranded Z-RNA through its unique Zα domain[94,95]. The ability of the Zα domain to make sequence-independent interactions with the Z-form phosphate backbone of both DNA and RNA enables ADAR1 to sense nucleic acid secondary structure conformations. Thus, double-stranded nucleotide-binding DRBPs can recognize their DNA and RNA targets through sequence-specific interactions (in the case of NF-κB) or through nonspecific interactions with the DNA and RNA backbone (in the case of ADAR1).

The evolution of DRBPs

The evolutionary forces driving the structure and function of DRBPs are complex, but understanding them will help us to identify new DRBPs and perhaps predict their susceptibility to interactions with lncRNAs. Although the DRBPs identified in our analysis are members of many different structural classes, each with their own evolutionary history, we focus on members of two very dissimilar DRBP families: CSD-containing proteins and eukaryotic DNA methyltransferases. CSD-containing proteins, which are required to protect cells from low temperatures, are members of an ancient DRBP family that uses weak selection criteria to interact with nucleic acids and therefore intrinsically bind to both DNA and RNA. Members of the eukaryotic DNA methyltransferase family are DRBPs that have more recently evolved the ability to recognize both DNA and RNA: they preferentially interact with DNA, and only one family member (DNMT2) acquired the ability to bind and methylate tRNAs[96]. The discovery of a eukaryotic DNA methyltransferase with a predominant tRNA methylation activity and only a modest DNA methylation activity showcases how evolution can modify protein surfaces to create new functions of DRBPs.

The ancient CSD DRBPs. The CSD is one of the most ancient nucleic acid-binding domains found in bacteria, archaea and eukaryotes. All CSD-containing proteins bind DNA and RNA (see Supplementary information S1 (table)). In humans, there are several CSD-containing proteins, such as the three Y box-binding proteins, LIN28A and LIN28B (which are homologues of the *Caenorhabditis elegans* LIN-28), and CSD-containing protein E1 (CSDE1). Y box 1 (YB1; also known as YBX1) was originally named for its ability to bind and repress the Y box of major histocompatibility complex class II promoters[97]. YB1 also binds RNA, with roles in alternative splicing[98], translational control[99] and RNA stabilization[100]. In addition, YB1 binds to damaged DNA and is involved in the DNA-damage response[101–103] — it translocates to the nucleus following stresses, such as exposure to ultraviolet radiation[104,105].

In bacteria, CSDs exist in short proteins that contain one CSD with short flanking sequences. In *Escherichia coli*, there are nine such proteins (cold-shock protein A (CspA)–CspI), which are probably products of multiple gene duplication events[106]. Of these, CspA, CspB, CspG and CspI are induced by cold stress, with CspA constituting >10% of all protein synthesized during cold shock[107–110]. Simultaneous deletion of these four genes results in lack of *E. coli* colony formation at temperatures at or lower than 25 °C (REF. 111). CspD is induced by nutrient stress[112], but CspC and CspE are constitutively expressed at normal growth temperature[113]. Many (if not all) of the Csp proteins bind DNA and RNA[114,115] and have similar roles to those of the human CSD-containing proteins, including in maintaining RNA stability[114], in translational regulation[116], in transcriptional control[116,117], in DNA replication and repair[118,119], and in chromosome folding[120].

CSD-containing proteins are widespread in plants[121], in which they have similar cellular functions. The first Csp-like protein found in plants was wheat CSP 1 (WCSP1), which is upregulated specifically by cold stress and binds ssDNA, dsDNA and RNA homopolymers[122]. WCSP1 was found to complement the cold-sensitive phenotype of the *E. coli* four-gene knockouts mentioned above[123], exhibiting remarkable functional conservation. In addition, WCSP1 showed nucleic acid melting activity in *E. coli*, which is critical to preventing inappropriate nucleic acid secondary structures that disrupt and terminate transcription. This activity is similar to the endogenous *E. coli* CspA, which also has transcription antitermination activity[123]. *Arabidopsis thaliana* has four CSD-containing proteins, CSP1–CSP4, all of which can also complement the quadruple *csp* knockout in *E. coli* to varying degrees, suggesting that their DNA and RNA interactions are well conserved during evolution[124–126].

Aptamers
Single-stranded DNA or RNA molecules that selectively bind small molecules, proteins and peptides with high affinity. Aptamers have dynamic tertiary structures, which contribute to the diversity of their binding targets.

Z-DNA
DNA in the conformation of a left-handed double helix.

Z-RNA
RNA in the conformation of a left-handed double helix.

Y box
A DNA target sequence with the consensus CTGATTG, which is recognized and bound by certain proteins.

RNA homopolymers
Sequences of ribonucleotides consisting of a single base; for example, CCCCCCCCC.

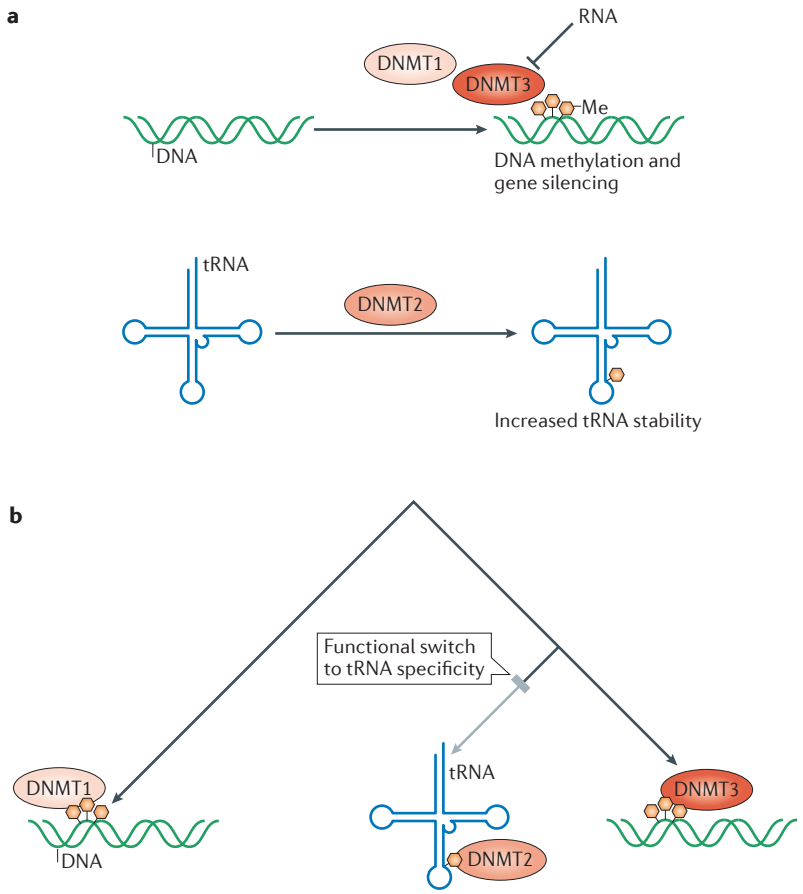

Figure 5 | **DNA methyltransferases target both DNA and RNA.** **a** | Best known for their role in gene silencing, all DNA methyltransferase family members are able to interact with both RNA and DNA[21,22,96]. DNA (cytosine-5)-methyltransferase 1 (DMNT1) and DMNT3 play a part in initiating and maintaining DNA methylation, whereas DNMT2 methylates tRNAs. This modification is critical for maintaining tRNA stability and cell viability. **b** | The evolution of the three major DNMTs[131] is depicted in a cladogram. DNMT2 probably diverged from its ancestral DNA methyltransferase to perform a critical role in methylating tRNAs, a function which it performs redundantly with NSUN2 (REF. 132). This radical change in substrate specificity highlights the ability of evolution to reshape a DNA-binding interface into one that preferentially recognizes RNA.

Unlike their counterparts in bacteria, most plant and animal CSD-containing proteins have additional functional domains (including more CSDs), which expand their functions, protein–protein interactions and/or nucleic acid-binding specificities. For example, the human protein CSDE1 has five CSDs, which increase the protein's affinity for target RNA sequences[127]. YB1 has both an N-terminal and a C-terminal domain flanking its CSD, which can support homomultimerization and interactions with many other protein partners (reviewed in REF. 128). In addition to its CSD, WCSP1 has three CCHC zinc-fingers, through which most of its dsDNA binding is mediated[122]. Nevertheless, the exceptional sequence and functional conservation between eukaryotic CSD-containing proteins and bacterial Csp proteins demonstrate a conserved, ancient role and origin of the domain. It is likely that a CSD fold that was capable of binding DNA and RNA was present in the last common ancestor of bacteria, archaea and eukaryotes[129].

The curious case of DNMT2. DNA methylation has important roles in gene expression and repression of transposable elements in eukaryotic cells. There are three eukaryotic proteins in the cytosine-C5 DNA methyltransferase family: DNMT1, DNMT2 and DNMT3. Whereas DNMT1 and DNMT3 play important parts in maintaining genome-wide methylation, DNMT2 has little DNA methylation activity[130] and is instead capable of methylating tRNAs[96] (FIG. 5a). When this activity was discovered, it was speculated that the three eukaryotic DNMTs might have evolved from an RNA methyltransferase[96]. However, there is no evidence that DNMT2 is more closely related to the ancestral protein of the family members. In fact, the three eukaryotic DNMTs may not be monophyletic and may have evolved from separate bacterial DNA methylation restriction-modification enzymes[131]. Thus, it seems likely that DNMT2 shifted its nucleic acid specificity from DNA to RNA in the last common eukaryotic ancestor[131] (FIG. 5b).

Despite the relatively narrow substrate specificity of DNMT2 compared with that of its family members, it is highly conserved and is the only extant DNMT in some species, such as *Schizosaccharomyces pombe* and *Drosophila melanogaster*[131]. This seems to indicate that DNMT2 has an important physiological role; however, *Dnmt2*$^{-/-}$ mice are viable and fertile, and yield no obvious phenotype[96]. This apparent contradiction was resolved with the recent report that deletion of DNMT2 in addition to another tRNA methyltransferase, NSUN2, in mice is lethal[132]. These mice show defects in tRNA stability, protein synthesis and differentiation[132], implying that the DNA methylation activity of DNMT2 is dispensable, whereas its tRNA methylation activity is not.

NSUN2 is a member of the nuclear protein 1 (NCL1) family of eukaryotic RNA cytosine-C5 methyltransferases, which are broadly distributed among eukaryotes[133]. Interestingly, NSUN2 itself is a DRBP and is able to bind and methylate both tRNA and hemi-methylated DNA[134]. Crosslinking immunoprecipitation-based analyses showed that NSUN2 also methylates mRNAs and non-coding RNAs[135]. Given the distant evolutionary relationship between DNA and RNA cytosine-C5 methyltransferases[131], NSUN2 and DNMT2 have most likely undergone convergent evolution from an RNA-binding and a DNA-binding protein family, respectively, to ensure proper tRNA modification. These evolutionary trajectories have bestowed on both proteins the ability (if residual) to bind and modify both DNA and RNA. This indicates not only that proteins with evolutionarily conserved DNA-binding activities are capable of binding RNA (and vice versa) but also that some nucleic acid substrates may be similar enough in sequence and structure to promote binding promiscuity. As mentioned above, this phenomenon is exploited by RNAs, both endogenous and artificial, that function as decoys to modulate DRBP function[1,19,136].

Conclusion and perspectives

In this Analysis, we have demonstrated that DRBPs constitute a significant fraction of cellular proteins — perhaps 2% of the human proteome — and have important

Intrinsically disordered protein
A protein that does not have a well-ordered three-dimensional structure, such as proteins containing random coils or multiple domains connected with flexible linkers.

Systematic evolution of ligands by exponential enrichment
(SELEX). A technique to identify ligand-binding-sequence specificity that is based on sequential rounds of binding of an oligonucleotide library to the ligand, followed by PCR amplification of the bound sequences.

cellular roles. Their functions include the control of transcription and translation, DNA repair, mediating responses to stress, splicing and apoptosis. These functions are intimately linked to the structures of DRBPs: orthogonal binding of DNA and RNA provides an opportunity for competitive regulation of transcription by decoy RNAs, whereas simultaneous binding of DNA and RNA permits transcriptional activation by RNA co-activators or allows the recruitment of RNA-containing complexes to specific DNA loci. In turn, the structures underlying DRBP functions are linked to their evolution. Some DRBPs contain ancient domains that have long bound DNA or RNA; others contain multiple domains that separately confer DNA- and RNA-binding abilities and mediate their functional roles.

The majority of RNA-binding proteins have had remarkably similar motifs during evolution[137], although individual members of protein families, such as the forkhead box transcription factors, can have diverse nucleic acid sequence specificities arising from independent evolutionary events[138]. It is also worth noting that intrinsically disordered protein domains that do not fold into defined secondary structures may also play important parts in mediating nucleic acid binding[14], as was found for RNA chaperones[139]. In addition to protein evolution, nucleic acid sequence evolution has important roles in the development of DRBP function. The discovery of lncRNAs illuminated new cellular binding targets for proteins that were previously thought of as DNA-specific binding proteins. Tens of thousands of human lncRNAs have been catalogued[140], and it is likely that many of them have yet-undiscovered functions requiring binding to proteins that are currently considered as DNA-specific binding proteins or that have so far only been shown to bind RNA *in vitro*. For example, the GCR and the oestrogen receptor were shown to bind DNA and RNA competitively >20 years before a physiological role for RNA–steroid receptor interactions was established[141–143]. Experimental selection techniques, such as systematic evolution of ligands by exponential enrichment (SELEX), have been used to develop inhibitory RNA aptamers for DNA-binding proteins, such as NF-κB. If such inhibitory RNA binding is functionally advantageous, the rapidly evolving sequences of lncRNAs[144] could provide a platform for the evolution of an analogous endogenous function, and many DRBPs may have species-specific RNA targets. For example, the RNA *Lethe*, which binds NF-κB, exists only in mice and is not present even in the closely related rat genome[145].

Proteins that bind both DNA and RNA could have several obvious functional advantages. By binding to both mRNAs and their encoding promoters, DRBPs can exert a powerful, amplified effect on gene expression. This also allows greater flexibility in generating cellular responses, as these DRBPs could both produce rapid effects on protein synthesis and impart long-acting changes on gene expression. At a cellular level, using one DRBP rather than two independent DNA-binding and RNA-binding proteins is more efficient, as it requires the transcription and translation of only one gene product. Finally, competitive RNA and DNA binding by some DRBPs allows an additional level of transcription factor regulation through decoy RNAs. These functional advantages, in addition to the rapid pace at which lncRNAs and their functions are being discovered, strongly indicate that more DRBPs and DRBP-mediated functions will be discovered in the coming years.

1. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA *GAS5* is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3**, ra8 (2010). **This paper shows that a widely expressed lncRNA was able to accumulate during stress and act as a decoy RNA to prevent steroid receptors from binding to their target DNA.**
2. Ishmael, F. T. *et al.* The human glucocorticoid receptor as an RNA-binding protein: global analysis of glucocorticoid receptor-associated transcripts and identification of a target RNA motif. *J. Immunol.* **186**, 1189–1198 (2011). **This study shows that the GCR binds to mRNAs involved in inflammation to accelerate their degradation.**
3. Davis, B. N., Hilyard, A. C., Nguyen, P. H., Lagna, G. & Hata, A. SMAD proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. *Mol. Cell* **39**, 373–384 (2010).
4. Shi, L. *et al.* Dynamic binding of Ku80, Ku70 and NF90 to the IL-2 promoter *in vivo* in activated T-cells. *Nucleic Acids Res.* **35**, 2302–2310 (2007).
5. Shi, L., Godfrey, W. R., Lin, J., Zhao, G. & Kao, P. N. NF90 regulates inducible IL-2 gene expression in T cells. *J. Exp. Med.* **204**, 971–977 (2007).
6. Xu, B. *et al.* Dax-1 and steroid receptor RNA activator (*SRA*) function as transcriptional coactivators for steroidogenic factor 1 in steroidogenesis. *Mol. Cell. Biol.* **29**, 1719–1734 (2009).
7. Lanz, R. B. *et al.* A steroid receptor coactivator, *SRA*, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17–27 (1999). **This paper details the discovery of a lncRNA involved in coordinating steroid receptor-driven gene activation.**
8. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genet.* **43**, 621–629 (2011). **This study describes a high-throughput approach to identify lncRNAs that interfere with transcription factors to repress gene expression, suggesting potentially widespread roles for promoter lncRNAs in cell-growth control.**
9. Rapicavoli, N. A. *et al.* A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife* **2**, e00762 (2013). **This study shows that the expression of pseudogene lncRNAs is actively regulated and that these lncRNAs may interact directly with proteins to modulate cell signalling.**
10. Lebruska, L. L. & Maher, L. J. 3rd. Selection and characterization of an RNA decoy for transcription factor NF-κB. *Biochemistry* **38**, 3168–3174 (1999).
11. Duss, O. *et al.* Structural basis of the non-coding RNA *RsmZ* acting as a protein sponge. *Nature* **509**, 588–592 (2014). **This paper shows how a bacterial lncRNA acts as a 'sponge' to sequester and release proteins involved in translation.**
12. Binns, D. *et al.* QuickGO: a web-based tool for gene ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
13. Hu, S. *et al.* Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* **139**, 610–622 (2009). **This paper details the use of a combined bioinformatics- and protein microarray-based strategy to systematically characterize the human protein–dsDNA interactome.**
14. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012). **This study identifies the mRNA interactome by using high-throughput covalent capture techniques coupled with proteomic-based detection strategies.**
15. Kadmiel, M. & Cidlowski, J. A. Glucocorticoid receptor signaling in health and disease. *Trends Pharmacol. Sci.* **34**, 518–530 (2013).
16. Reddy, T. E. *et al.* Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res.* **19**, 2163–2171 (2009).
17. Clark, A. R. & Belvisi, M. G. Maps and legends: The quest for dissociated ligands of the glucocorticoid receptor. *Pharmacol. Ther.* **134**, 54–67 (2012).
18. Kanai, A., Tani, H., Torimura, M. & Akimitsu, N. The RNA degradation pathway regulates the function of *GAS5* a non-coding RNA in mammalian cells. *PLoS ONE* **8**, e55684 (2013).
19. Fukunaga, J. *et al.* The Runt domain of AML1 (RUNX1) binds a sequence-conserved RNA motif that mimics a DNA element. *RNA* **19**, 927–936 (2013).
20. Gros, C. *et al.* DNA methylation inhibitors in cancer: recent and future approaches. *Biochimie* **94**, 2280–2296 (2012).
21. Holz-Schietinger, C. & Reich, N. O. RNA modulation of the human DNA methyltransferase 3A. *Nucleic Acids Res.* **40**, 8550–8557 (2012).
22. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013). **This paper describes how RNA transcripts are able to regulate local DNA methylation by interacting with DNA methyltransferases.**
23. Cattaneo, A., Biocca, S., Corvaja, N. & Calissano, P. Nuclear localization of a lactic dehydrogenase with single-stranded DNA-binding properties. *Exp. Cell Res.* **161**, 130–140 (1985).
24. Calissano, P., Volontè, C., Biocca, S. & Cattaneo, A. Synthesis and content of a DNA-binding protein with lactic dehydrogenase activity are reduced by nerve growth factor in the neoplastic cell line PC12. *Exp. Cell Res.* **161**, 117–129 (1985).

25. Pioli, P. A., Hamilton, B. J., Connolly, J. E., Brewer, G. & Rigby, W. F. Lactate dehydrogenase is an AU-rich element-binding protein that directly interacts with AUF1. *J. Biol. Chem.* **277**, 35738–35745 (2002).
26. Demarse, N. A. *et al.* Direct binding of glyceraldehyde 3-phosphate dehydrogenase to telomeric DNA protects telomeres against chemotherapy-induced rapid degradation. *J. Mol. Biol.* **394**, 789–803 (2009).
27. Singh, R. & Green, M. Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. *Science* **259**, 365–368 (1993).
This study identifies GAPDH as a tRNA-binding protein, demonstrating in principle how tRNA levels may control metabolism and how metabolites may influence protein synthesis.
28. Nagy, E. & Rigby, W. F. C. Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD^+-binding region (Rossmann fold). *J. Biol. Chem.* **270**, 2755–2763 (1995).
29. Ray, R. & Miller, D. M. Cloning and characterization of a human c-Myc promoter-binding protein. *Mol. Cell. Biol.* **11**, 2154–2161 (1991).
30. Hernandez-Perez, L. *et al.* α-enolase binds to RNA. *Biochimie* **93**, 1520–1528 (2011).
31. Fletcher, L., Rider, C. C. & Taylor, C. B. Enolase isoenzymes. III. Chromatographic and immunological characteristics of rat brain enolase. *Biochim. Biophys. Acta* **452**, 245–252 (1976).
32. Kato, K. *et al.* Immunoassay of human muscle enolase subunit in serum: a novel marker antigen for muscle diseases. *Clin. Chim. Acta* **131**, 75–85 (1983).
33. Anderson, S. L., Minard, K. I. & McAlister-Henn, L. Allosteric inhibition of NAD^+-specific isocitrate dehydrogenase by a mitochondrial mRNA. *Biochemistry* **39**, 5623–5629 (2000).
34. Hentze, M. W. & Preiss, T. The REM phase of gene regulation. *Trends Biochem. Sci.* **35**, 423–426 (2010).
35. Auphan, N., DiDonato, J. A., Rosette, C., Helmberg, A. & Karin, M. Immunosuppression by glucocorticoids: inhibition of NF-κB activity through induction of IκB synthesis. *Science* **270**, 286–290 (1995).
36. Surjit, M. *et al.* Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell* **145**, 224–241 (2011).
37. Jonat, C. *et al.* Antitumor promotion and antiinflammation: down-modulation of AP-1 (Fos/Jun) activity by glucocorticoid hormone. *Cell* **62**, 1189–1204 (1990).
38. Dhawan, L., Liu, B., Blaxall, B. C. & Taubman, M. B. A novel role for the glucocorticoid receptor in the regulation of monocyte chemoattractant protein-1 mRNA stability. *J. Biol. Chem.* **282**, 10146–10152 (2007).
39. Yang-Yen, H.-F. *et al.* Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein–protein interaction. *Cell* **62**, 1205–1215 (1990).
40. Ray, A. & Prefontaine, K. E. Physical association and functional antagonism between the p65 subunit of transcription factor NF-κB and the glucocorticoid receptor. *Proc. Natl Acad. Sci. USA* **91**, 752–756 (1994).
41. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **2008**, 102–114 (2008).
42. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nature Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
43. Massagué, J. & Wotton, D. Transcriptional control by the TGF-β/SMAD signaling system. *EMBO J.* **19**, 1745–1754 (2000).
44. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61 (2008).
45. Kumarswamy, R., Volkmann, I. & Thum, T. Regulation and function of miRNA-21 in health and disease. *RNA Biol.* **8**, 706–713 (2011).
46. Kao, P. N. *et al.* Cloning and expression of cyclosporin A- and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. *J. Biol. Chem.* **269**, 20691–20699 (1994).
47. Kuwano, Y. *et al.* NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. *Nucleic Acids Res.* **38**, 225–238 (2009).
48. Shim, J., Lim, H., R. Yates Iii, J. & Karin, M. Nuclear export of NF90 is required for interleukin-2 mRNA stabilization. *Mol. Cell* **10**, 1331–1344 (2002).
49. Sakamoto, S. *et al.* The NF90–NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol. Cell. Biol.* **29**, 3754–3769 (2009).
50. Malek, T. R. & Castro, I. Interleukin-2 receptor signaling: at the interface between tolerance and immunity. *Immunity* **33**, 153–165 (2010).
51. Dienz, O. & Rincon, M. The effects of IL-6 on CD4 T cell responses. *Clin. Immunol.* **130**, 27–33 (2009).
52. Iliopoulos, D., Hirsch, H. A. & Struhl, K. An epigenetic switch involving NF-κB, lin28, let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell* **139**, 693–706 (2009).
53. Takamizawa, J. *et al.* Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.* **64**, 3753–3756 (2004).
54. Guo, N. L. *et al.* Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. *Clin. Cancer Res.* **14**, 8213–8220 (2008).
55. Ng, S.-Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The long noncoding RNA *RMST* interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359 (2013).
56. Sarkar, A. & Hochedlinger, K. The Sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15–30 (2013).
57. Ng, S.-Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* **31**, 522–533 (2012).
References 55 and 57 describe how the lncRNA *RMST* physically interacts with the transcription factor SOX2 to co-regulate a large pool of downstream genes implicated in neurogenesis.
58. Feng, J. *et al.* The *Evf-2* noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470–1484 (2006).
59. Bond, A. M. *et al.* Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature Neurosci.* **12**, 1020–1027 (2009).
60. Jeffery, L. & Nakielny, S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J. Biol. Chem.* **279**, 49479–49487 (2004).
61. Wassenegger, M., Heimes, S., Riedel, L. & Sänger, H. L. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).
62. Schmitz, K. M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269 (2010).
63. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289–1292 (2004).
64. Sun, B. K., Deaton, A. M. & Lee, J. T. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol. Cell* **21**, 617–628 (2006).
65. Zazopoulos, E., Lalli, E., Stocco, D. M. & Sassone-Corsi, P. DNA binding and transcriptional repression by DAX-1 blocks steroidogenesis. *Nature* **390**, 311–315 (1997).
66. Kabe, Y. *et al.* The role of human MBF1 as a transcriptional coactivator. *J. Biol. Chem.* **274**, 34196–34202 (1999).
67. Xu, B. & Koenig, R. J. An RNA-binding domain in the thyroid hormone receptor enhances transcriptional activation. *J. Biol. Chem.* **279**, 33051–33056 (2004).
68. Zhao, X. *et al.* Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell* **15**, 549–558 (2004).
69. Watanabe, M. *et al.* A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor α coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, *SRA*. *EMBO J.* **20**, 1341–1352 (2001).
70. Deblois, G. & Giguere, V. Ligand-independent coactivation of ERα AF-1 by steroid receptor RNA activator (*SRA*) via MAPK activation. *J. Steroid Biochem. Mol. Biol.* **85**, 123–131 (2003).
71. Poon, M. M. & Chen, L. Retinoic acid-gated sequence-specific translational control by RARα. *Proc. Natl Acad. Sci. USA* **105**, 20303–20308 (2008).
72. van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
73. Court, R., Chapman, L., Fairall, L. & Rhodes, D. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep.* **6**, 39–45 (2005).
74. Chesnokov, I. N. in *International Review of Cytolology. Vol. 256* (ed. Kwang, W. J.) 69–109 (Academic Press, 2007).
75. Deng, Z., Dheekollu, J., Broccoli, D., Dutta, A. & Lieberman, P. M. The origin recognition complex localizes to telomere repeats and prevents telomere-circle formation. *Curr. Biol.* **17**, 1989–1995 (2007).
76. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. *TERRA* RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35**, 403–413 (2009).
77. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc.* **4**, 44–57 (2009).
78. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
79. Maris, C., Dominguez, C. & Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131 (2005).
80. Fiset, S. & Chabot, B. hnRNP A1 may interact simultaneously with telomeric DNA and the human telomerase RNA *in vitro*. *Nucleic Acids Res.* **29**, 2268–2275 (2001).
81. Bahadur, R. P., Zacharias, M. & Janin, J. Dissecting protein–RNA recognition sites. *Nucleic Acids Res.* **36**, 2705–2716 (2008).
82. Allers, J. & Shamoo, Y. Structure-based analysis of protein–RNA interactions using the program ENTANGLE. *J. Mol. Biol.* **311**, 75–86 (2001).
83. Buratti, E. *et al.* Nuclear factor TDP43 and SR proteins promote *in vitro* and *in vivo CFTR* exon 9 skipping. *EMBO J.* **20**, 1774–1784 (2001).
84. Kawahara, Y. & Mieda-Sato, A. TDP43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc. Natl Acad. Sci. USA* **109**, 3347–3352 (2012).
85. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP43. *Nature Struct. Mol. Biol.* **20**, 1443–1449 (2013).
86. Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J. & Yuan, H. S. Structural insights into TDP43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37**, 1799–1808 (2009).
87. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845 (1999).
88. Hayden, M. S. & Ghosh, S. NF-κB, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* **26**, 203–234 (2012).
89. Wurster, S. E. & Maher, L. J. Selection and characterization of anti-NF-κB p65 RNA aptamers. *RNA* **14**, 1037–1047 (2008).
90. Huang, D. B. *et al.* Crystal structure of NF-κB $(p50)_2$ complexed to a high-affinity RNA aptamer. *Proc. Natl Acad. Sci. USA* **100**, 9268–9273 (2003).
91. Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. Structure of the NF-κB p50 homodimer bound to DNA. *Nature* **373**, 311–317 (1995).
92. Lu, X.-J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protoc.* **3**, 1213–1227 (2008).
93. Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L. & Breaker, R. R. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11**, 774–784 (2005).
94. Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A. & Rich, A. Crystal structure of the Zα domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841–1845 (1999).
95. Placido, D., Brown, B. A., Lowenhaupt, K., Rich, A. & Athanasiadis, A. A left-handed RNA double helix bound by the Zα domain of the RNA-editing enzyme ADAR1. *Structure* **15**, 395–404 (2007).
96. Goll, M. G. *et al.* Methylation of $tRNA^{Asp}$ by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395–398 (2006).
This study shows that the DNA methylase DNMT2 evolved the ability to methylate tRNA.

97. Didier, D. K., Schiffenbauer, J., Woulfe, S. L., Zacheis, M. & Schwartz, B. D. Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc. Natl Acad. Sci. USA* **85**, 7322–7326 (1988).
98. Stickeler, E. *et al.* The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *EMBO J.* **20**, 3821–3830 (2001).
99. Skabkina, O. V., Lyabin, D. N., Skabkin, M. A. & Ovchinnikov, L. P. YB-1 autoregulates translation of its own mRNA at or prior to the step of 40S ribosomal subunit joining. *Mol. Cell. Biol.* **25**, 3317–3323 (2005).
100. Evdokimova, V. *et al.* The major mRNA-associated protein YB-1 is a potent 5′ cap-dependent mRNA stabilizer. *EMBO J.* **20**, 5491–5502 (2001).
101. Marenstein, D. R. *et al.* Stimulation of human endonuclease III by Y box-binding protein 1 (DNA-binding protein B). Interaction between a base excision repair enzyme and a transcription factor. *J. Biol. Chem.* **276**, 21242–21249 (2001).
102. de Souza-Pinto, N. C. *et al.* Novel DNA mismatch-repair activity involving YB-1 in human mitochondria. *DNA Repair* **8**, 704–719 (2009).
103. Ise, T. *et al.* Transcription factor Y-box binding protein 1 binds preferentially to cisplatin-modified DNA and interacts with proliferating cell nuclear antigen. *Cancer Res.* **59**, 342–346 (1999).
104. Stein, U. *et al.* Hyperthermia-induced nuclear translocation of transcription factor YB-1 leads to enhanced expression of multidrug resistance-related ABC transporters. *J. Biol. Chem.* **276**, 28562–28569 (2001).
105. Koike, K. *et al.* Nuclear translocation of the Y-box binding protein by ultraviolet irradiation. *FEBS Lett.* **417**, 390–394 (1997).
106. Yamanaka, K., Fang, L. & Inouye, M. The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Mol. Microbiol.* **27**, 247–255 (1998).
107. Etchegaray, J. P., Jones, P. G. & Inouye, M. Differential thermoregulation of two highly homologous cold-shock genes, *cspA* and *cspB*, of *Escherichia coli*. *Genes Cells* **1**, 171–178 (1996).
108. Wang, N., Yamanaka, K. & Inouye, M. CspI, the ninth member of the CspA family of *Escherichia coli*, is induced upon cold shock. *J. Bacteriol.* **181**, 1603–1609 (1999).
109. Nakashima, K., Kanamaru, K., Mizuno, T. & Horikoshi, K. A novel member of the *cspA* family of genes that is induced by cold shock in *Escherichia coli*. *J. Bacteriol.* **178**, 2994–2997 (1996).
110. Goldstein, J., Pollitt, N. S. & Inouye, M. Major cold shock protein of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **87**, 283–287 (1990).
111. Xia, B., Ke, H. & Inouye, M. Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in *Escherichia coli*. *Mol. Microbiol.* **40**, 179–188 (2001).
112. Yamanaka, K. & Inouye, M. Growth-phase-dependent expression of *cspD*, encoding a member of the CspA family in *Escherichia coli*. *J. Bacteriol.* **179**, 5126–5130 (1997).
113. Phadtare, S. & Inouye, M. Role of CspC and CspE in regulation of expression of RpoS and UspA, the stress response proteins in *Escherichia coli*. *J. Bacteriol.* **183**, 1205–1214 (2001).
114. Jiang, W., Hou, Y. & Inouye, M. CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J. Biol. Chem.* **272**, 196–202 (1997).
115. Phadtare, S. & Inouye, M. Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of *Escherichia coli*. *Mol. Endocrinol.* **33**, 1004–1014 (1999).
116. Hanna, M. M. & Liu, K. Nascent RNA in transcription complexes interacts with CspE, a small protein in *E. coli* implicated in chromatin condensation. *J. Mol. Biol.* **282**, 227–239 (1998).
117. Bae, W., Xia, B., Inouye, M. & Severinov, K. *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. *Proc. Natl Acad. Sci. USA* **97**, 7784–7789 (2000).
118. Yamanaka, K., Zheng, W., Crooke, E., Wang, Y. H. & Inouye, M. CspD, a novel DNA replication inhibitor induced during the stationary phase in *Escherichia coli*. *Mol. Microbiol.* **39**, 1572–1584 (2001).
119. Chen, C. S. *et al.* A proteome ChIP approach reveals new DNA damage recognition activities in *Escherichia coli*. *Nature Methods* **5**, 69–74 (2008).
120. Hu, K. H. *et al.* Overproduction of three genes leads to camphor resistance and chromosome condensation in *Escherichia coli*. *Genetics* **143**, 1521–1532 (1996).
121. Karlson, D. & Imai, R. Conservation of the cold shock domain protein family in plants. *Plant Physiol.* **131**, 12–15 (2003).
122. Karlson, D., Nakaminami, K., Toyomasu, T. & Imai, R. A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. *J. Biol. Chem.* **277**, 35248–35256 (2002).
123. Nakaminami, K., Karlson, D. T. & Imai, R. Functional conservation of cold shock domains in bacteria and higher plants. *Proc. Natl Acad. Sci. USA* **103**, 10122–10127 (2006).
124. Imai, R., Kim, M. H., Sasaki, K., Sato, S. & Sonoda, Y. *Plant and Microbe Adaptations to Cold in a Changing World* (Springer, 2013).
125. Juntawong, P., Sorenson, R. & Bailey-Serres, J. Cold shock protein 1 chaperones mRNAs during translation in *Arabidopsis thaliana*. *Plant J.* **74**, 1016–1028 (2013).
126. Kim, J. S. *et al.* Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **35**, 506–516 (2007).
127. Triqueneaux, G., Velten, M., Franzon, P., Dautry, F. & Jacquemin-Sablon, H. RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Res.* **27**, 1926–1934 (1999).
128. Eliseeva, I. A., Kim, E. R., Guryanov, S. G., Ovchinnikov, L. P. & Lyabin, D. N. Y-box-binding protein 1 (YB-1) and its functions. *Biochem.* **76**, 1402–1433 (2012).
129. Mihailovich, M., Militti, C., Gabaldón, T. & Gebauer, F. Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays* **32**, 109–118 (2010).
130. Hermann, A., Schmitt, S. & Jeltsch, A. The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J. Biol. Chem.* **278**, 31717–31721 (2003).
131. Lyko, F., Jurkowski, T. P. & Jeltsch, A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS ONE* **6**, e28104 (2011).
132. Tuorto, F. *et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nature Struct. Mol. Biol.* **19**, 900–905 (2012).
133. Pavlopoulou, A. & Kossida, S. Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. *Genomics* **93**, 350–357 (2009).
134. Sakita-Suto, S. *et al.* Aurora-B regulates RNA methyltransferase NSUN2. *Mol. Biol. Cell* **18**, 1107–1117 (2007).
135. Hussain, S. *et al.* NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* **4**, 255–261 (2013).
136. Reiter, N. J., Maher, L. J. & Butcher, S. E. DNA mimicry by a high-affinity anti-NF-κB RNA aptamer. *Nucleic Acids Res.* **36**, 1227–1236 (2008).
This paper shows that the RNA aptamer developed in reference 10 mimics DNA to target the DNA-binding interface of NF-κB.
137. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
138. Nakagawa, S., Gisselbrecht, S. S., Rogers, J. M., Hartl, D. L. & Bulyk, M. L. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl Acad. Sci. USA* **110**, 12349–12354 (2013).
139. Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169–1175 (2004).
140. Volders, P. J. *et al.* LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* **41**, D246–D251 (2012).
141. Ali, M. & Vedeckis, W. The glucocorticoid receptor protein binds to transfer RNA. *Science* **235**, 467–470 (1987).
142. Feldman, M., Kallos, J. & Hollander, V. P. RNA inhibits estrogen receptor binding to DNA. *J. Biol. Chem.* **256**, 1145–1148 (1981).
143. Rossini, G. P. RNA-containing nuclear binding sites for glucocorticoid-receptor complexes. *Biochem. Biophys. Res. Commun.* **123**, 78–83 (1984).
144. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
References 2, 10, 57, 143 and 144 reveal that RNA binding to transcription factors may be a widespread phenomenon.
145. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755 (2013).

Acknowledgements
W.H.H. is supported by American Heart Association predoctoral fellowship 13PRE16920012 and by a US National Institutes of Health (NIH) training grant to Emory University (5T32GM008602-14). Work in the laboratory of E.A.O. is supported by grant R01DK095750 from the NIH National Institute of Diabetes and Digestive and Kidney Diseases and by grant 14GRNT20460124 from the American Heart Association.

Competing interests statement
The authors declare no competing interests.

DATABASE
QuickGo: http://www.ebi.ac.uk/QuickGO
GO:0003723 | GO:0003677

FURTHER INFORMATION
Unpublished structure of TDP43 RRMs: http://dx.doi.org/10.2210/pdb4iuf/pdb

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF